

Deduplication Engine

You can check for and run reports on the duplicate comments received on any document that has received at least 20 public submissions. The Deduplication engine helps agencies in the timely and efficient reviewing and processing of comments especially when they are received in large quantities.

Accessing the Deduplication Engine

(A) Click on the Document ID of the FR Document (not shown here).

Note: Document Details tab is displayed by default.

(B) On the **Document Details** tab, from the top right section of the screen, click on the **Actions** Dropdown Menu and select **Run Deduplication**.

Note: If the Deduplication action has already been run on the FR, the Actions Dropdown Menu will show **View Deduplication Results** option as well.

Note: The Assigned Docket Manager or an Agency Administrator is able to Run, Re-Run, or View the Deduplication results. All FDMS roles that have access to the FR Document can view the Deduplication results.

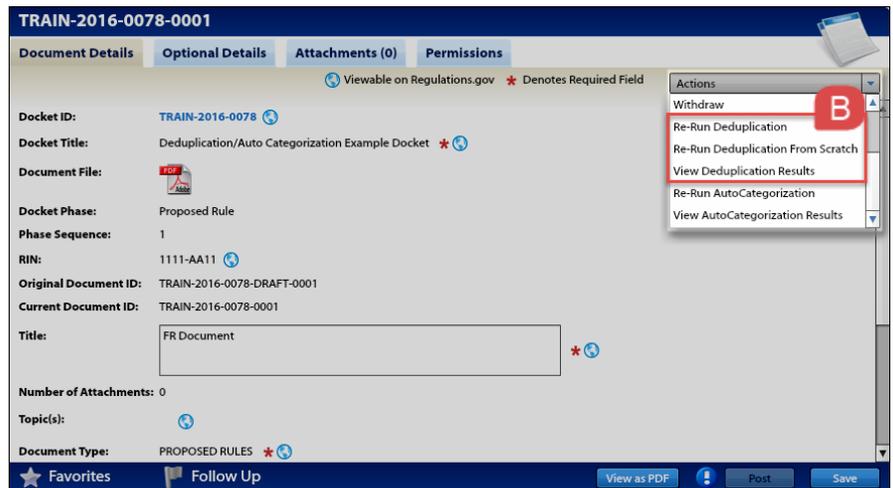


Figure 1: Accessing Deduplication tool

Deduplication Results

The **Deduplication Results** are displayed in a split screen. On the left is the **Duplicates Tree** and on the right is the Main Display Frame showing the **Deduplication Details** with a bar diagram.

(A) Each folder with the Document ID contains the Pivot and its near duplicate comments.

(B) At the bottom of the Duplicates Tree is the folder containing the **Unique** comments (not shown here).

(C) Click the **Re-Run Deduplication** button to rerun the deduplication.

(D) Click on a bar in the diagram to show the comments in that particular folder under the Duplicates Tree.

(E) View the **Deduplication Details** such as the FR Document ID, Title, Docket ID and the Number of Pivot, Duplicates and Unique Documents.

(F) Click on the **Generate Report** button to generate a report in CSV format and save it in your computer or external sources.

Note: When multiple documents are selected, the **Multiple Document Selection** screen appears (not shown here) with the batch processing options (refer to Batch Processing QRG for details).

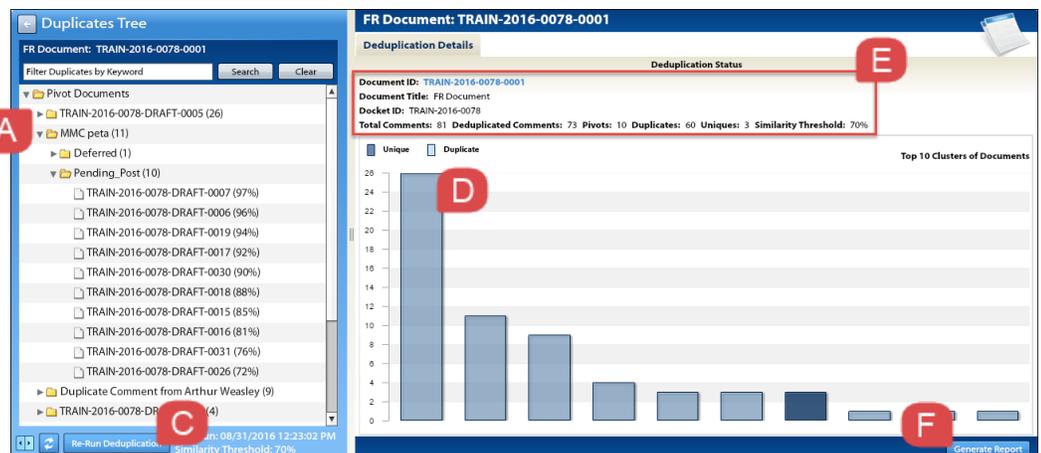


Figure 2: Deduplication Results

Reviewing near Duplicate Documents

- (A) Click on the Pivot document under the Duplicates Tree to view the Pivot document on the right (not shown here).
- (B) Select the near duplicate comment ID under the Pivot document to see the **Deduplication Details** on the right.
- (C) See the percentage similarity of the near duplicate comment with the Pivot in parenthesis at the end of the Document ID.
- (D) Navigate to the previous or the next comment by clicking on the respective buttons. You can also mark the comment as unique.
- (E) To display a compact view of the comment, uncheck the box next to Show Whitespace.

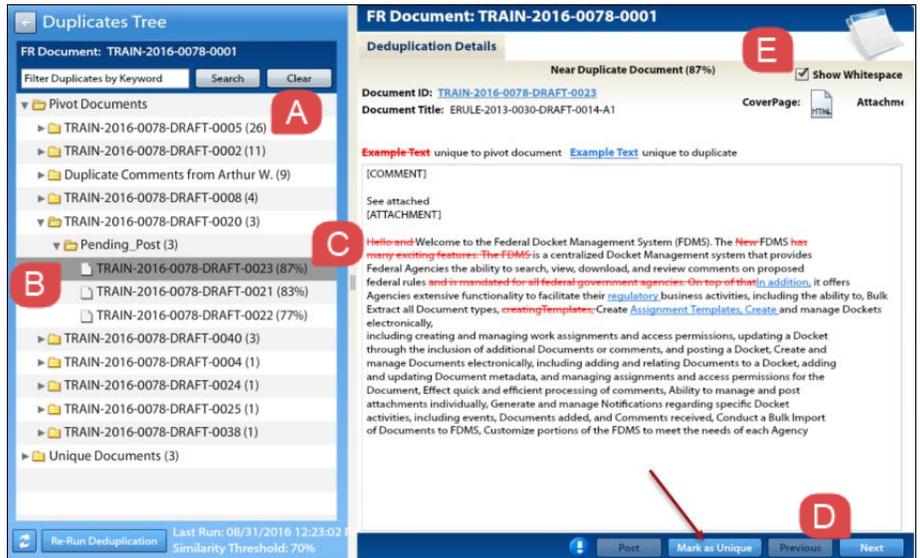


Figure 4: Near Duplicate Comment

- ❖ **Red strikethrough** = phrases from the Pivot Document that are not included in the near duplicate comment.
- ❖ **Blue underline** = additional phrases in the near duplicate comment that are not in the Pivot Document.
- ❖ **Black** = same phrases contained in both Pivot Document and near duplicate comment.

Deduplication Analysis Report

From the Deduplication Results screen,

- ❖ Click on the Generate Report button as shown in Figure 2.
- ❖ Select the appropriate program on the pop up screen to open the report (not shown here).
- ❖ Save the report on the desktop or other external sources for further customized analysis of deduplication data.

(A) Pivot Document is the comment identified by the engine as the baseline to detect the near duplicate comments.

(B) Near Duplicate Document is a comment that is similar to a specified level of threshold to the Pivot Document and is included in the set of near duplicate comments.

Document ID	Similarity	Pivot Document ID	Count of Documents in the Pivot	Count of Exact Duplicates of the Pivot
TRAIN-2016-0078-DRAFT-0002			11	0
TRAIN-2016-0078-DRAFT-0007	97	TRAIN-2016-0078-DRAFT-0002		
TRAIN-2016-0078-DRAFT-0006	96	TRAIN-2016-0078-DRAFT-0002		
TRAIN-2016-0078-DRAFT-0019	94	TRAIN-2016-0078-DRAFT-0002		
TRAIN-2016-0078-DRAFT-0017	92	TRAIN-2016-0078-DRAFT-0002		
TRAIN-2016-0078-DRAFT-0030	90	TRAIN-2016-0078-DRAFT-0002		
TRAIN-2016-0078-DRAFT-0018	88	TRAIN-2016-0078-DRAFT-0002		
TRAIN-2016-0078-DRAFT-0015	85	TRAIN-2016-0078-DRAFT-0002		
TRAIN-2016-0078-DRAFT-0016	81	TRAIN-2016-0078-DRAFT-0002		
TRAIN-2016-0078-DRAFT-0031	76	TRAIN-2016-0078-DRAFT-0002		
TRAIN-2016-0078-DRAFT-0026	72	TRAIN-2016-0078-DRAFT-0002		
TRAIN-2016-0078-DRAFT-0037	72	TRAIN-2016-0078-DRAFT-0002		
TRAIN-2016-0078-0006			9	9
TRAIN-2016-0078-DRAFT-0055	100	TRAIN-2016-0078-0006		
TRAIN-2016-0078-DRAFT-0056	100	TRAIN-2016-0078-0006		

Figure 5: Deduplication Analysis Report

- (C) Similarity column shows the percentage similarity of the near duplicate comment with the Pivot Document.
- (D) View the total number of near duplicate Documents in the Pivot folder.
- (E) View the **Count of Exact Duplicates of the Pivot** Document i.e. 100% similar.